Metrics for Benchmarking Computational Workload Reduction

Christoph Busch, Pawel Drozdowski,
Christian Rathgeb, Patrick Schuch

copy of slides available at:
https://www.christoph-busch.de/about-talks-slides.html

NIST-IFPC-2019
November 29, 2018

CRISP
Center for Research
in Security and Privacy

NTNU

Overview

Agenda

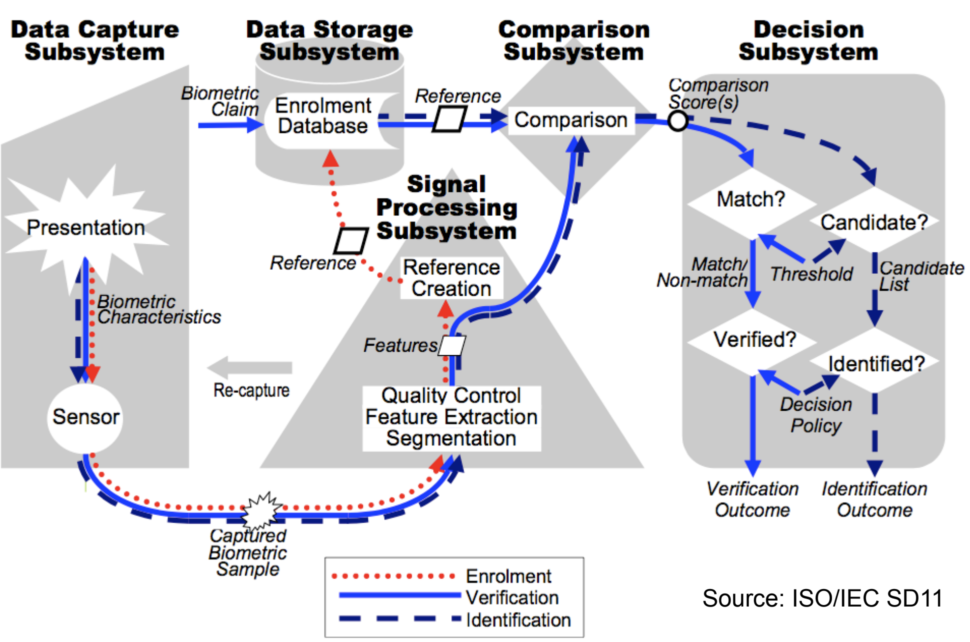
- Problem description
- Techniques for computational workload reduction
- Standardised metrics for identification systems evaluation
- Future - what should be added to ISO/IEC 19795-1
- Conclusion

Problem Description

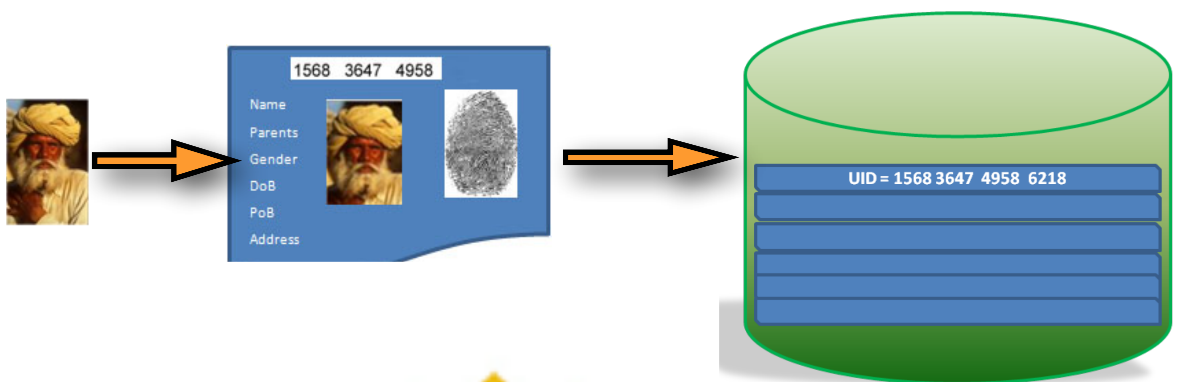
Diversity of Applications

Biometric applications (as defined in ISO/IEC 2382-37)

- **Verification**:
  process of confirming a biometric claim through biometric comparison
  - computational trivial case of a 1:1 comparison
- **Identification**:
  process of searching against a biometric enrolment database to find and return the biometric reference identifier(s) attributable to a single individual
  - in the worst case: compare a probe against all enrolled references



Source: ISO/IEC SD11



AADHAAR

Challenges of Identification Applications

Exhaustive search (naive approach)

- Increasing risk of false positive decision
  - The probability becomes quickly unacceptable:
    linear increase with size N of the database
  - This is expressed in ISO/IEC 19795-1:2006 with the
    FPIR definition in Clause 4.6.9. See:
    https://www.iso.org/obp/ui/#iso:std:iso-iec:19795:-1:ed-1:v1:en

- Increasing costs
  - Faced by large scale deployments (e.g. forensic systems)
  - Leading to upscaling of the infrastructure (hardware costs) and increasing operational costs (complexity of the infrastructure)
  - Leading to reduced usability (transaction time)
    for instance for mobile police personnel
    requesting response from centralized forensic system
  - Leading to delays in de-duplication tasks

Challenges of Identification Applications

Some examples of large databases

- single 1:1 transaction with COTS fingerprint system [Neu17]
- 1:N grows linearly , N:N grows quadratically

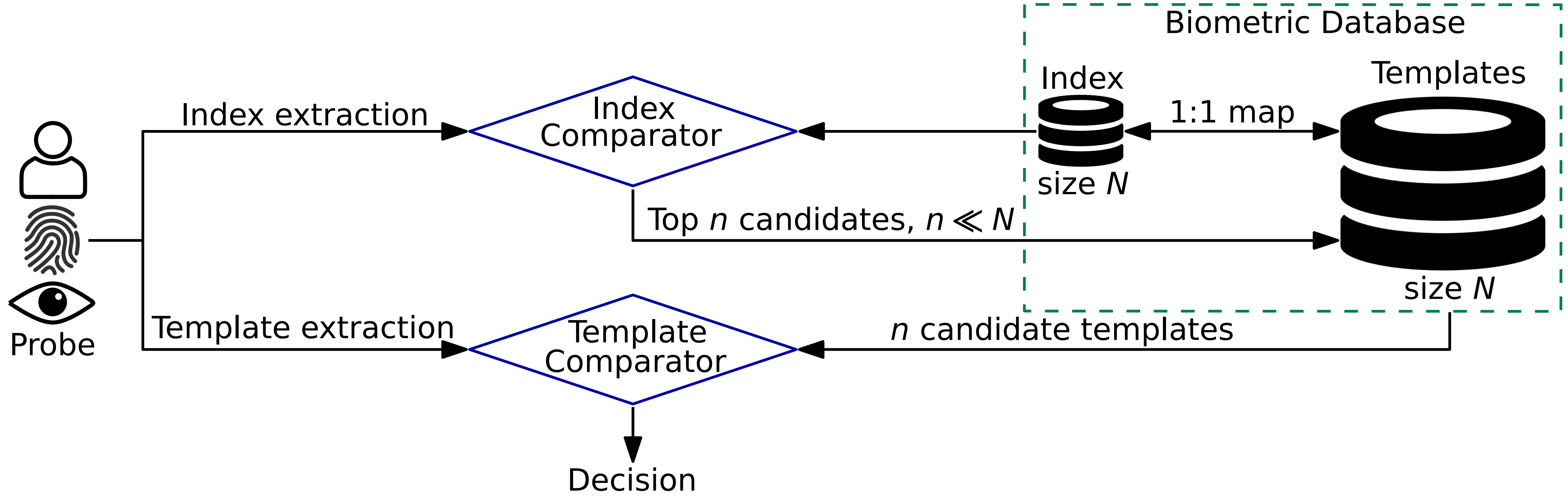
	NTNU Norwegian University of Science and Technology	Oslo	Norway	Schengen
Size	40k	600k	5M	500M
1:N	0.03s	0.5s	4s	7m
N:N	22m	3d	241d	>6000y

Techniques for Computational Workload Reduction - a.k.a as Indexing Methods

Workload Reduction - An Overview

Computational Workload Reduction Methods

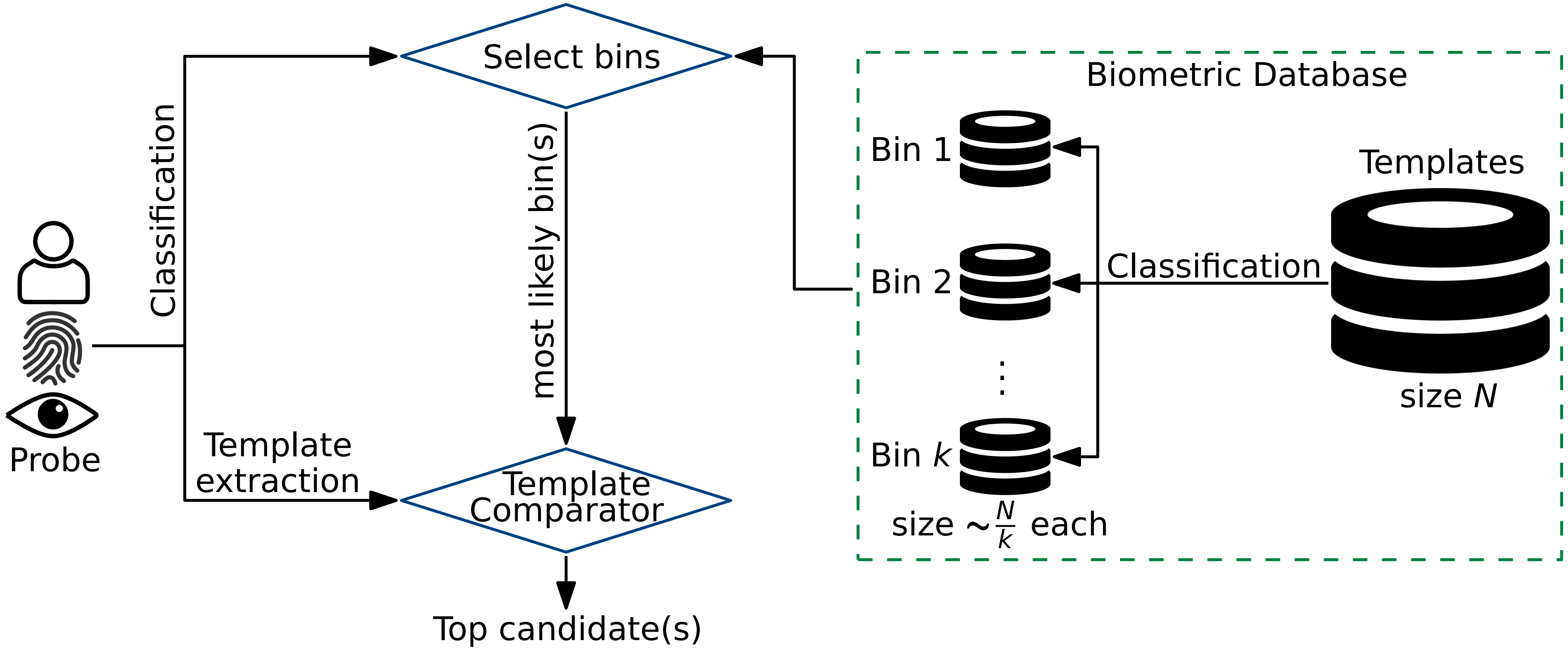
- Cascading algorithms, Serial combination and Pre-selection
  - The probe is exhaustively compared to the enrolled templates using a computationally efficient (but somewhat inaccurate) comparator/algorithm.
  - A candidate (short)list (significantly smaller than the whole DB) is produced.
  - The candidate (short)list is searched exhaustively using the normal, accurate (but computationally expensive) comparator/algorithm [Gent2009]



Penetration rate can be reduced

Computational Workload Reduction Methods (2)

- Binning, Classification, Clustering
  - The DB is split into a number of bins/classes/clusters (e.g. based on metadata like sex, ethnicity, age, or statistical features of the templates).
  - Template comparisons are performed within the bin/class/cluster of the DB corresponding to that of the probe [Mhatre2005]

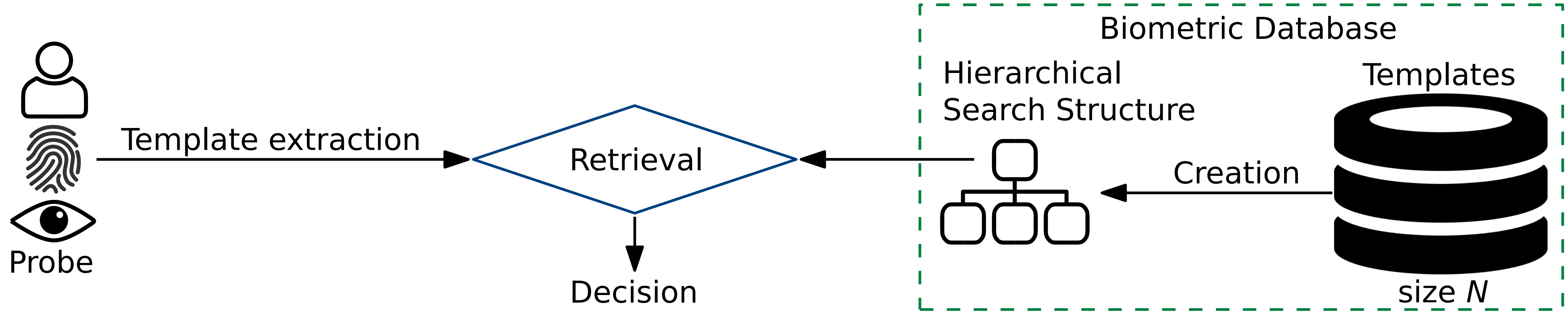


Penetration rate can be reduced

Workload Reduction - An Overview

Computational Workload Reduction Methods (3)

- Hierarchical retrieval
  - An efficient search structure (e.g. trees, fuzzy hashing) for the DB is created.
  - The retrieval of candidate list/identity proceeds in sub-linear time [Proenca2017]



Penetration rate can be reduced

Methods can be combined (e.g. binning followed by indexing)

Workload Reduction - An Overview

Computational Workload Reduction Methods (4)

- Efficient representations
  - The size or form of templates is changed (e.g. through binarisation) thus making them more compact or capable of utilising more efficient instructions, particularly the bitwise operators [Xu2008]
  - Other properties of templates are changed (e.g. exhibiting pose/alignment invariance, and thus not needing to compensate for those during comparisons, for instance as is the case for Iris-Codes and circular shifting) [Rathg2013]
- Efficient comparators
  - The comparator is augmented in some way (e.g. by taking advantage of some intrinsic template properties), thus requiring less computational workload [Rathg2016]

Computational cost of single template comparison can be reduced

Workload Reduction - An Overview

Relevance in system evaluations
(see FRVT-2018 presentation by Patrick Grother)

NIST

FR Performance Revolution

» Error rates dropped by order-of-magnitude last 4-5 years
- Implications for marketplace "tech refresh" ⟶ Now!
- Reputational benefit for face recognition
- Implications for demographic differentials: ΔAB reduced

» Industry expansion
- 20 developers better now than then industry leader NEC was in 2014
- But... large variation across industry ⟶ buyer beware.

» Template sizes have contracted, vary across industry
- Leading algorithms: 256 – 4200 bytes, most accurate 1024 bytes

» Search durations reduced also
- Search durations x10000 across industry
- Most linear, but some sublinear, approaching logarithmic

Source: P. Grother (NIST) - report on FRVT-2018

An evaluation shall report accuracy AND search duration

Standardised Metrics for Identification System Evaluation

Which Metrics do we have today ?

Metrics to evaluate identification systems
are defined in ISO/IEC 19795-1:2006

- Accuracy determined by recognition performance
  - **false-positive identification-error rate (FPIR)**
    proportion of identification transactions by users not enrolled in the system, where an identifier is returned
  - **false-negative identification-error rate (FNIR)**
    proportion of identification transactions by users enrolled in the system in which the user's correct identifier is not among those returned
- Search duration only indicated by penetration rate and pre-selection error (p-s-e rate is the complement to the hit rate)
  - **penetration rate**
    <pre-selection algorithm> measure of the average number of pre-selected templates as a fraction of the total number of templates
  - If binning/classification/clustering is in place, then we report the
    pre-selection error rate
    proportion of genuine attempts where the enrolment template corresponding to the input sample is not in the pre-selected subset of templates
    that would be compared with the input sample

Why is this not sufficient?

What we should add to Standard-Methodology

As we can combine multiple computational workload reduction methods

- the pure penetration rate is not sufficient to report about duration
- computational workload can be reduced irrespective of the penetration rate (e.g. different, more efficient template representations in an exhaustive search)

Duration of a single transaction depends on

- number of enrolled references (# of data subjects in the DB)
- computational workload (i.e. of the transaction in the biometric system under test)
  - workload is dependent on hardware (processor and memory available) on which the system is operating
  - this is not necessarily reproducible by another testing lab
- which workload reduction methods are combined

Therefore: for a given hardware environment (SOTA baseline)

we need to measure workload reduction in terms of

- workload difference (w.r.t. to the selected baseline) @ defined number of enrolled references

What we should add to Standard-Methodology

ISO/IEC 19795-1:202x is currently under revision

- The 3rd Working Draft (WD) is
  - Waiting for comments by 2018-11-30
  - Containing a definition in Clause 4.29 for
    computational workload
    total computational effort of a single transaction (or set of transactions) in a biometric system, including execution time, memory requirements, etc.
  - Indicating in Clause 8.10.2, what must be considered for identification systems
    - *Generation of a biometric probe from the captured biometric sample*
    - *Pre-selection to reduce workload of identification search*
    - *Identification search over the reference database*
    - *Production of candidate list and deciding identification outcome*

The proposed metrics should be **hardware independent**, if possible.

Therefore the **number of intrinsic operations** is more relevant than execution time: For example the number of bit or float comparisons will allow a cross-platform benchmark.

What we should add to Standard-Methodology

ISO/IEC 19795-1:202x is currently under revision

- The next Working Draft (WD) should also contain a new metric in Clause 8.10.2 for
  - **computational workload (CW)**
    - which considers the number of enrolees N
    - the penetration rate p
    - the cost of a single feature vector comparison C
    - the cost of the pre-selection c
    - the costs for production of the candidate list and decision l

$$CW = N * p * C + c + l$$

The cost for pre-processing (e.g. segmentation) is negligible, as it is conducted for the probe only.

What we should add to Standard-Methodology

ISO/IEC 19795-1:202x is currently under revision

- Then we have the illustrating new metric in Clause 8.10.3 for
  - **computational workload difference (CWD)**
    - which is the proportion of workload w.r.t. to the baseline system (SOTA)
    - tested on a select hardware
    - takes into account the number N of enrolees

$$CWD(N) = 1 - \frac{CW_i}{CW_b}$$

- where CW_i is the $i\text{-}th$ system under test
- where CW_b is the baseline system chosen by the evaluator

We subtract the fraction of the computational workload reduction from the baseline, which is 1 or 100%

Example Evaluation

Example Evaluation

According to the proposed metric

- Suppose an iris identification system with N = 1000 enrollees and for the sake of simplicity assume the decision costs (l) such as candidate list sorting to be negligibly small.
- In the baseline scenario, a state-of-the-art iris-code based system is used with:
  - Template size of 10.240 bits
  - Hamming distance based comparator performing 17 circular shifts for alignment compensation, i.e. C = 10.240 * 17 = 174.080 bit comparisons
  - Exhaustive search (p = 1.0, c = 0.0)
- Further, suppose a system with a pre-selection algorithm [Gent2009], where computationally efficient templates are used in the first step to create a candidate shortlist, followed by the aforementioned state-of-the-art algorithm in the second step operating on the shortlist only:
  - 5% of the original database size is pre-selected as a candidate shortlist, i.e. p = 0.05
  - The compact templates have the size of 2048 bits, are compared using Hamming distance, and require no alignment compensation. Hence, the pre-selection costs are: c = 1000 * 2048 = 2.048 * 10^6 bit comparisons

Example Evaluation

According to the proposed metric

- The computational workload of this baseline is then:
  - $CW_b = 1000 * 1.0 * 174.080 + 0 = 1.7408 * 10^8$ bit comparisons
- The computational workload of the system is then:
  - $CW_i = 1000 * 0.05 * 174.080 + 2.048 * 10^6 = 1.0752 * 10^7$ bit comparisons
- Finally, the computational workload difference between the proposed system and a state-of-the-art baseline at 1000 enrollees is:
  - $CWD(1000) = 1 - (1.0752 * 10^7 / 1.7408 * 10^8) = 93.82\%$
  - in other words, the proposed system reduces the computational workload by over 90% w.r.t. the baseline system

Future - What needs to be done?

Conclusion

In order to learn, where and how to improve identification systems, we need

- to measure computational workload reduction in terms of transaction duration
- and combine accuracy testing reports with duration testing reports

Future work

- There are numerous competitions on this topic, which should be aligned to a standardised metric, e.g.
  - Bologna: FIDX-TEST
    https://biolab.csr.unibo.it/fvcongoing/UI/Form/ICB2013FIDX.aspx
    https://biolab.csr.unibo.it/FvcOnGoing/UI/Form/PublishedAlgs.aspx
  - NIST: FRVT 1:N 2018 Evaluation
    https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt-1n-2018-evaluation

References

Publications

- [Gent2009] J. E. Gentile, N. Ratha, and J. Connell: „An efficient, two-stage iris recognition system“, in Intl. Conf. on Biometrics: Theory, Applications and Systems (BTAS), pages 211–215. IEEE, (2009)
- [Mahtre2005] A. J. Mhatre, S. Palla, S. Chikkerur, and V. Govindaraju: „Efficient search and retrieval in biometric databases“, in Biometric technology for human identification II, volume 5779, pages 265–274. International Society for Optics and Photonics, (2005)
- [Proenca2017] H. Proenca and J. C. Neves: „Iris biometric indexing“, pages 101–124, Security. Institution of Engineering and Technology, (2017)
- [Xu2008] H. Xu, R. N. J. Veldhuis, T. A. M. Kevenaar, A. H. M. Akkermans, and A. M. Bazen: „Spectral minutiae: A fixed-length representation of a minutiae set“, in Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, (2008)
- [Rathg2013] C. Rathgeb, F. Breitinger, and C. Busch: „Alignment-free cancelable iris biometric templates based on adaptive Bloom filters“, in Intl. Conf. on Biometrics (ICB), pages 1–8. IEEE, (2013)
- [Rathg2016] C. Rathgeb, H. Hofbauer, A. Uhl, and C. Busch: „TripleA: Accelerated accuracy-preserving alignment for iris-codes“, in Intl. Conf. on Biometrics (ICB), pages 1–8. IEEE, (2016)
- [ProGuiti2014] P. D. Gutierrez, M. Lastra, F. Herrera, and J. M. Benıtez: „A high performance fingerprint matching system for large databases based on GPU“, in Trans. on Information Forensics and Security, 9(1):62–71, (2014)
- [Drod2019] P. Drozdowski, C. Rathgeb, C. Busch: „Computational Workload in Large-Scale Biometric Identification Systems“,
  under review

Contact

Prof. Dr. Christoph Busch

Norwegian University of Science and Technology
Department of Information Security and Communication Technology
Teknologiveien 22
2802 Gjøvik, Norway
Email: `christoph.busch@ntnu.no`
Phone: +47-611-35-194

Contact

CRISP
Center for Research
in Security and Privacy

h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Prof. Dr. Christoph Busch
Principal Investigator

Hochschule Darmstadt FBI
Haardtring 100
64295 Darmstadt, Germany
christoph.busch@crisp-da.de

Telefon +49-6151-16-30090
www.dasec.h-da.de
www.crisp-da.de